

User-centered Evaluation of Adaptive and Adaptable Systems

Lex S. van Velsen, Thea M. van der Geest, Rob F. Klaassen

University of Twente

Faculty of Behavioural sciences, Institute for Behavioural Research

Department of Technical and Professional Communication

P.O. Box 217, 7500 AE Enschede, the Netherlands

{l.s.vanvelsen, t.m.vandergeest, r.f.klaassen}@gw.utwente.nl

Abstract. Adaptive and adaptable systems provide tailored output to various users in various contexts. While adaptive systems base their output on implicit inferences, adaptable systems use explicitly provided information. Since the presentation or output of these systems is adapted, standard user-centered evaluation methods do not produce results that can be easily generalized. This calls for a reflection on the appropriateness of standard evaluation methods for user-centered evaluations of these systems. We have conducted a literature review to create an overview of the methods that have been used. When reviewing the empirical evaluation studies we have, among other things, focused on the variables measured and the implementation of results in the (re)design process. The goal of our review has been to compose a framework for user-centered evaluation. In the next phase of the project, we intend to test some of the most valid and feasible methods with an adaptive or adaptable system.

1 Background

In this working paper our ongoing study into the user-centered evaluation of adaptable and adaptive systems will be discussed. First we will introduce our definitions and the issues concerning user-centered evaluation of adaptive and adaptable systems. Subsequently we deal with the research questions and the method used for generating a literature overview and framework for user-centered evaluation. Last, we discuss some preliminary results of our literature review.

We use the following definitions for adaptive and adaptable systems in our study, based on Benyon et al. [1].

Adaptive systems can alter aspects of their structure, functionality or interface on the basis of a user model generated from implicit user input, in order to accommodate the differing needs of individuals or groups of users and the changing needs of users over time.

Adaptable systems do not create a user model based on implicit inferences from the interaction between the user and the system. They explicitly ask for input.

Adaptable systems can alter aspects of their structure, functionality or interface on the basis of a user model generated from explicit user input, in order to accommodate the differing needs of individuals or groups of users and the changing needs of users over time.

Many times, a system is not just adaptive or adaptable, but uses a combination of both techniques [2]. For a definition that addresses the possible appearance of both techniques in a single system, we use the term ‘personalized systems’.

Personalized systems can alter aspects of their structure, functionality or interface on the basis of a user model generated from implicit and / or explicit user input, in order to accommodate the differing needs of individuals or groups of users and the changing needs of users over time.

Our definition of user-centered evaluation stems from the user-centered design principles by Gould and Lewis [3] and the definition of user-centered approach as can be found in ISO 13407 [4].

User-centered evaluation deals with the empirical evaluation of a system by gathering subjective user feedback on satisfaction and productivity, quality of work, support and training costs and user health and well-being.

2 Personalized Systems and User-Centered Evaluation

The output or appearance of a personalized system differs for every individual user or group of users in every context. This adapted output has the potential to be a great benefit for users: it is geared towards the user’s preferences, behavior or needs and can make interaction easier and more fruitful.

The system’s adapted output calls for a reconsideration of the appropriateness of user-centered evaluation methods when applied to personalized systems. Many evaluation methods build on the assumption that the output of a system is the same for every user, but this assumption does not hold for personalized systems. Generalization of results can be problematic. Do results gathered from a few subjects hold for an entire population when every person is presented with different system output? Moreover, new usability criteria must be applied to personalized systems [5]. Traditional user-centered evaluation methods do not take these criteria into account. The evaluation that is the subject of our study can be classified as the third and fourth layer of evaluation according to Weibelzahl’s categorization [6] and deals with the subjective appreciation of the adaptation decision and the total interaction. Well functioning of preceding layers is a prerequisite for a successful evaluation of layer number three and four.

Reported empirical evaluations of personalized systems are relatively scarce [7], and little is written about the characteristics of the user-centered evaluation method applied to the personalized characteristics of the system. As a result it is unclear which user-centered evaluation method one should choose for the evaluation of a certain system. Furthermore, implications of the possible methods are not documented completely, which makes it difficult for a researcher to anticipate appropriately for the trade-offs that every evaluation method causes.

Our study has two goals. The first goal is to provide an overview of user-centered evaluation methods that can be applied to personalized systems. This takes shape in a literature review of empirical user-centered evaluations of personalized systems which will focus on each method's characteristics and (dis)advantages. A framework for the user-centered evaluation of personalized systems will be the intended yield of this review. The second goal of our study is to expand the body of knowledge concerning user-centered methodologies for the evaluation of personalized systems. This will be done by testing one or two methods from the literature review, which scored high on validity and feasibility.

The literature review to be included in our study will be discussed in more detail in this working paper.

3 Research Questions for the Literature Review

For the literature review of reports on user-centered evaluation of personalized systems, the following main question is formulated:

Which user-centered evaluation methods are suitable and feasible for evaluating personalized systems?

This main question is accompanied by six secondary questions that address relevant variables of each evaluation method:

- Which types of personalized systems have been evaluated?
- Which methods were used for the user-centered evaluation of personalized systems?
- With regard to the methods used for the user-centered evaluation of personalized systems, what is / are its: (dis)advantages, validity, reliability and costs?
- What kind of participants / test subjects were involved in the user-centered evaluation of personalized systems?
- Which variables have been measured in the user-centered evaluation of personalized systems?
- How have the results of user-centered evaluation of personalized systems been implemented in further (re)design processes?

4 Literature Review Approach

For our literature review we used the York method [8]. In order to find relevant studies, we searched eight databases with a range of search terms that could result in relevant hits. Databases ranged from the more technical databases as INSPEC to the more psychological PsycInfo. Publication lists of researchers active in the field of adaptive or adaptable systems and the reports included in the Easy-D database were also screened for relevance.

To narrow down the huge amount of potentially relevant reports of evaluations found after the search, selection criteria were applied. These criteria specify that a study has to report empirical data concerning the evaluation of a system that is adaptive and / or adaptable. At least some part of the evaluation should be user-centered and address at least one of the points included in the secondary questions. Evaluations that only discussed algorithms or purely technical evaluations were excluded from the review. Finally, studies should be reported in English, German or Dutch, and published after 1989. This year was chosen to ensure a big number of evaluations, at the same time excluding obsolete techniques.

Categorizations of methods applied and variables measured were deductively generated from the collected data. These categorizations and other variables were recorded in a database which will be published on a website when the study is completed. Full results will be published in an article which will include an overview of variables and corresponding methods.

5 Preliminary Results

Over 4,000 possibly relevant titles were scanned, resulting in 338 read abstracts, 127 articles fully read and 59 articles included in the review. Of the included articles, 37% dealt with adaptive systems, 25% with adaptable systems and 38% with systems that incorporated both techniques.

The evaluated systems were mostly learning systems (17%) and intelligent tourist guides (14%). Questionnaires (73%), interviews and data logging (both 42%) were the methods applied most. Methods almost never stand alone, but are part of an 'evaluation suite', containing several methods. Variables measured most are usability (53%), perceived usefulness of a system (41%) and user behaviour (39%).

The reporting of evaluations can be improved on and reflection on the evaluation process is scarce. There are different interpretations of (user-centered) evaluation to be found in the literature. Many times 'evaluation' consists of technical evaluation of the system and if the role of the user is involved, it is usually small (see for a typical example [9]).

Feasibility is an important attribute of a user-centered evaluation method because it often restrains the thoroughness of evaluations. Budgets are small and time is limited so evaluations have to be quick and cheap, if conducted at all. A full scale simulation of a mall as in [10] ensures a very good evaluation, but is costly. Testing the prototype of personalized information systems takes time when one wants it to be done properly because the database has to be 'fed' with data first [11].

The utilization of questionnaires sometimes leads to problems with regard to the validity of the evaluation method. Causes are for example: constructs compiled of only one or two items [12] or a small sample size (e.g., 11 subjects in [13]).

The key to improved evaluations lies with the authors reporting user-centered evaluations. When evaluation processes are documented properly and supplemented with a reflection on the evaluation in the discussion section, it is possible to learn not only about the system but also about the evaluation itself.

References

1. Benyon, D.R., Innocent, P.R. & Murray, D.M.: System Adaptivity and the Modeling of Stereotypes. Paper Presented at INTERACT '87, Second IFIP Conference on Human-Computer Interaction, the Netherlands (1987)
2. Wu, D., Im, I., Tremaine, M., Instone, K. & Turoff, M.: A Framework of Classifying Personalization Scheme used on E-commerce Websites. Paper Presented at the 36th Hawaii International Conference on System Sciences (2003)
3. Gould, J.D. & Lewis, C.: Designing for Usability: Key Principles and What Designers Think. *Communications of the ACM* **28** (1985) 300-311
4. International Standard Organization: ISO 13407: Human-centred Design Processes for Interactive Systems (1999)
5. Jameson, A.: Adaptive Interfaces and Agents. In: Jacko, J.A. & Sears, A. (eds.): *Human-Computer Interaction Handbook*. Mahwah, NJ, Erlbaum (2003) 305-330
6. Weibelzahl, S.: Evaluation of Adaptive Systems. Freiburg: Pedagogical University (2003)
7. Chin, D.N.: Empirical Evaluation of User Models and User-Adapted Systems. *User Modeling and User-Adapted Interaction* **11** (2001) 181-194
8. NHS Centre for Reviews and Dissemination: Undertaking Systematic Reviews of Research on Effectiveness. York: University of York (2001)
9. Díaz, A., Gervas, P. & García, A.: Evaluation of a System for Personalized Summarization of Web Contents. In: Ardissono, L., Brna, P. & Mitrovic, A. (eds.): *UM 2005, LNAI 3538* (2005) 453-462
10. Bohnenberger, T., Jameson, A., Krüger, A. & Butz, A.: Location-Aware Shopping Assistance: Evaluation of a Decision-Theoretic Approach. Paper Presented at the 4th International Symposium on Human-Computer Interaction with Mobile Devices. Pisa, Italy (2002)
11. de Roure, D., Hall, W., Reich, S., Hill, G., Pikrakis, A., Stairmand, M.: MEMOIR – an Open Framework for Enhanced Navigation of Distributed Information. *Information Processing and Management* **37** (2001) 53-74
12. Henderson, R., Rickwood, D. & Roberts, P.: The Beta Test of an Electronic Supermarket. *Interacting with Computers* **10** (1998) 385-399
13. de Almeida, P. & Yokoi, S.: Interactive Character as a Virtual Tour Guide to an Online Museum Exhibition. Paper Presented at Museums and the Web (2003)